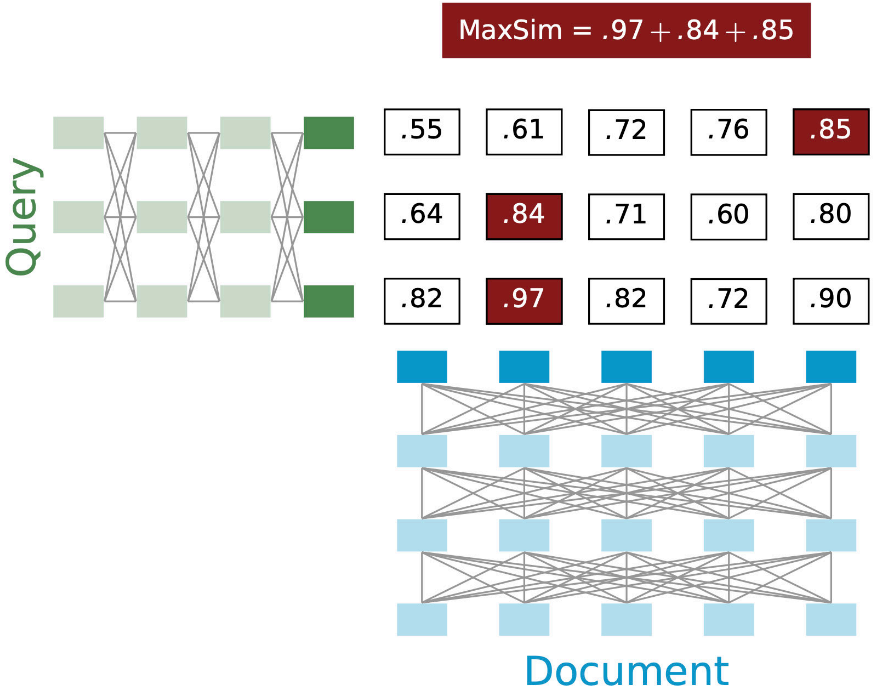
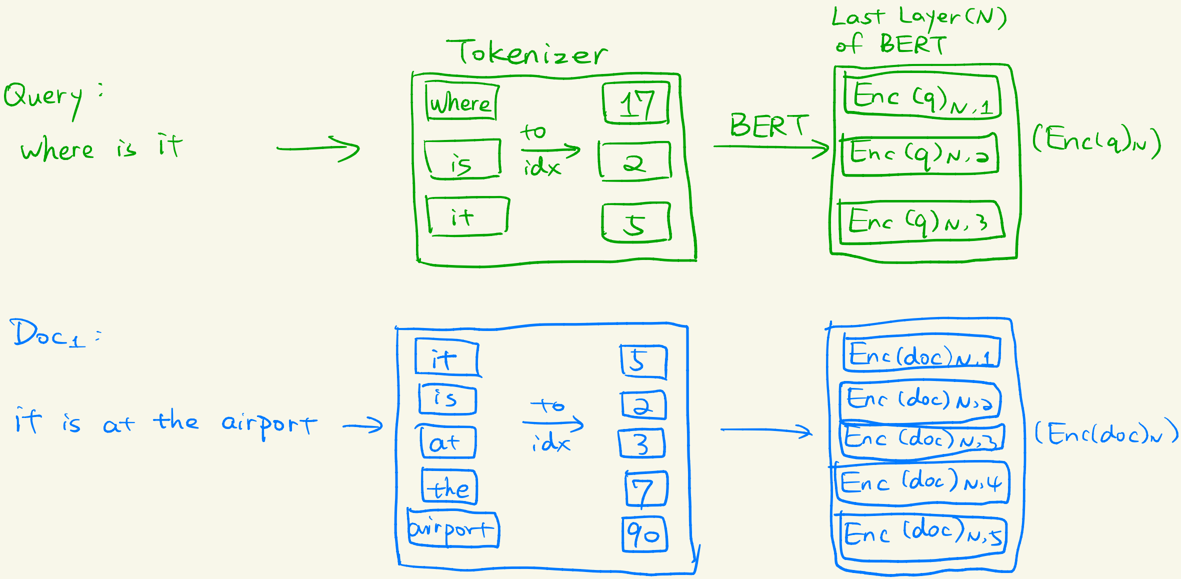


Notice that ColBERT represents the document as a MATRIX, not a vector like the bi-encoder! $\mathbf{d_i} \in \mathbb{R}^{L \times d}$

Query : where is it

Doc$_1$: it is at the airport

To Embeddings



In PA3, $Enc(q)_{N,i} \in \mathbb{R}^{728}$ $\Rightarrow$ $Enc(q)_N \in \mathbb{R}^{3\times 728}$

$Enc(doc)_{N,i} \in \mathbb{R}^{728}$ $\Rightarrow$ $Enc(doc)_N \in \mathbb{R}^{5\times 728}$ for this example

ColBERT (Zero-shot Transfer Ver.)

For convenience, we denote $E_q^i := \mathrm{Enc}(q)_{N,i}$, $E_d^i := \mathrm{Enc}(doc)_{N,i}$

(here we didn't apply L2 norm to the embeddings, but you are welcome to do so)

Query Embedding (where is it)

$$
\begin{array}{c|ccccc}
\boxed{E_q^1} & \boxed{E_q^{1\top} E_d^1} & \boxed{\boxed{E_q^{1\top} E_d^2}} & - & - & \boxed{E_q^{1\top} E_d^5} \\
\boxed{E_q^2} & \boxed{E_q^{2\top} E_d^1} & \boxed{E_q^{2\top} E_d^2} & - & - & \boxed{\boxed{E_q^{2\top} E_d^5}} \\
\boxed{E_q^3} & \boxed{\boxed{E_q^{3\top} E_d^1}} & \boxed{E_q^{3\top} E_d^2} & - & - & \boxed{E_q^{3\top} E_d^5} \\
\hline
 & \boxed{E_d^1} & \boxed{E_d^2} & \boxed{E_d^3} & \boxed{E_d^4} & \boxed{E_d^5}
\end{array}
$$

$E_q^i \in \mathbb{R}^{728}$

$E_d^i \in \mathbb{R}^{728}$

$\therefore E_q^{i\top} \cdot E_d^i \in \mathbb{R}$

Doc 1 Embedding (it is at the airport)

Assume the scores being framed are the maximum across each row.

Then the similarity for this (Query, Doc) pair will be

$$
\mathrm{MaxSim}(q, \mathrm{doc}_1) = \sum_i^L \max_j^M \mathrm{Enc}(q)_{N,i}{}^\top \mathrm{Enc}(\mathrm{doc})_{N,j} = \sum_i^3 \max_j^5 E_q^{i\top} \cdot E_d^j
$$

$$
= E_q^{1\top} \cdot E_d^2 + E_q^{2\top} E_d^5 + E_q^{3\top} \cdot E_d^1
$$

where L is the length of q, M is the length of Doc.

Then calculate $MaxSim(q, doc_i)$ for each of the 15 docs retrieved, and get the top-5 docs.